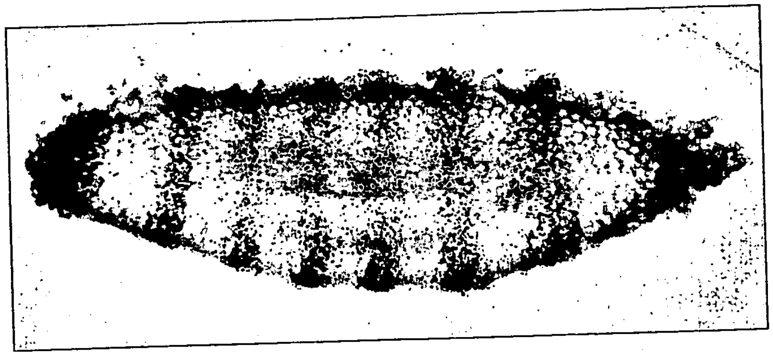
Exhibit 3

Textbook chapter from
Molecular Biology of the Cell, 3rd edition,
Garland Publishing,Inc. New York (1994)

MOLECULAR BIOLOGY OF
THE CELL
THIRD EDITION

Bruce Alberts • Dennis Bray
Julian Lewis • Martin Raff • Keith Roberts
James D. Watson

Garland Publishing, Inc.
New York & London



100 µm

Figure 7–20 ***In situ* hybridization for RNA localization in tissues.** Autoradiograph of a section of a very young *Drosophila* embryo that has been subjected to *in situ* hybridization using a radioactive DNA probe complementary to a gene involved in segment development. The probe has hybridized to RNA in the embryo, and the pattern of autoradiographic silver grains reveals that the RNA made by the gene (called *ftz*) is localized in alternating stripes across the embryo that are three or four cells wide. At this stage of development (cellular blastoderm), the embryo contains about 6000 cells. (From E. Hafen, A. Kuriowa, and W.J. Gehring, *Cell* 37:833–841, 1984. © Cell Press.)

a complementary partner strand, even in a cell or cell extract that contains millions of different DNA and RNA sequences. Probes of this type are widely used to detect the nucleic acids corresponding to specific genes, both to facilitate the purification and characterization of the genes after cell lysis and to localize them in cells, tissues, and organisms. Moreover, by carrying out hybridization reactions under conditions of "reduced stringency," a probe prepared from one gene can be used to find its evolutionary relatives—both in the same organism, where the relatives form part of a gene family, and in other organisms, where the evolutionary history of the nucleotide sequence can be traced.

DNA Cloning [15]

In **DNA cloning,** a DNA fragment that contains a gene of interest is inserted into the purified DNA genome of a self-replicating genetic element—generally a virus or a plasmid. A DNA fragment containing a human gene, for example, can be joined in a test tube to the chromosome of a bacterial virus, and the new *recombinant DNA molecule* can then be introduced into a bacterial cell. Starting with only one such recombinant DNA molecule that infects a single cell, the normal replication mechanisms of the virus can produce more than 10^{12} identical virus DNA molecules in less than a day, thereby amplifying the amount of the inserted human DNA fragment by the same factor. A virus or plasmid used in this way is known as a *cloning vector,* and the DNA propagated by insertion into it is said to have been *cloned.*

A DNA Library Can Be Made Using Either Viral or Plasmid Vectors [16]

In order to clone a specific gene, one begins by constructing a *DNA library*—a comprehensive collection of cloned DNA fragments, including (one hopes) at least one fragment that contains the gene of interest. The library can be constructed using either a virus or a plasmid vector and is generally housed in a population of bacterial cells. The principles underlying the methods used for cloning genes are the same for either type of cloning vector, although the details may be different. For simplicity, in this chapter we ignore these differences and illustrate the methods with reference to plasmid vectors.

The **plasmid vectors** used for gene cloning are small circular molecules of double-stranded DNA derived from larger plasmids that occur naturally in bacterial cells. They generally account for only a minor fraction of the total host bacterial cell DNA, but they can easily be separated on the basis of their small size from chromosomal DNA molecules, which are large and precipitate as a pellet upon centrifugation. For use as cloning vectors, the purified plasmid DNA circles are first cut with a restriction nuclease to create linear DNA molecules. The cel-

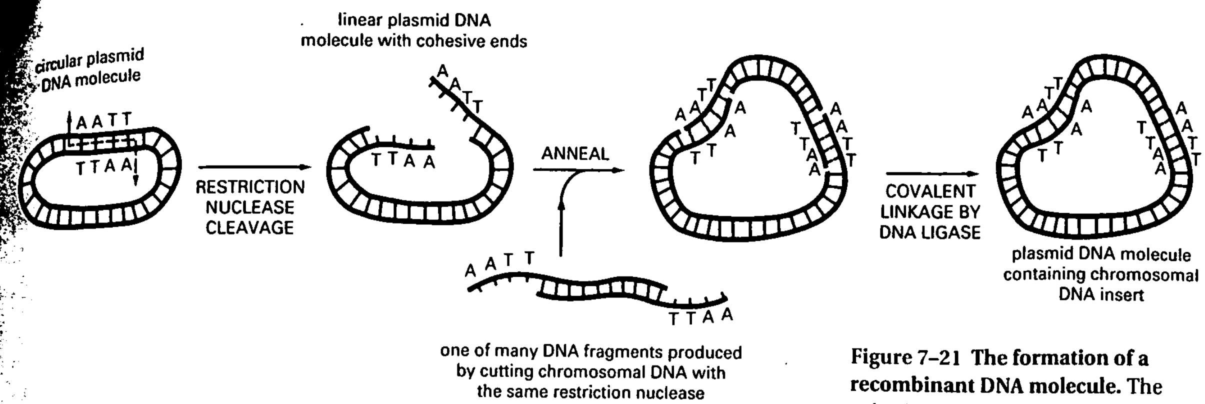


Figure 7–21 The formation of a recombinant DNA molecule. The cohesive ends produced by many kinds of restriction nucleases allow two DNA fragments to join by complementary base-pairing (see Figure 7–2). DNA fragments joined in this way can be covalently linked in a highly efficient reaction catalyzed by the enzyme DNA ligase. In this example a recombinant plasmid DNA molecule containing a chromosomal DNA insert is formed.

lular DNA to be used in constructing the library is cut with the same restriction nuclease, and the resulting restriction fragments (including those containing the gene to be cloned) are then added to the cut plasmids and annealed via their cohesive ends to form recombinant DNA circles. These recombinant molecules containing foreign DNA inserts are then covalently sealed with the enzyme DNA ligase (Figure 7–21).

In the next step in preparing the library, the recombinant DNA circles are introduced into bacterial cells that have been made transiently permeable to DNA; such cells are said to be *transfected* with the plasmids. As these cells grow and divide, doubling in number every 30 minutes, the recombinant plasmids also replicate to produce an enormous number of copies of DNA circles containing the foreign DNA (Figure 7–22). Many bacterial plasmids carry genes for antibiotic resistance, a property that can be exploited to select those cells that have been successfully transfected; if the bacteria are grown in the presence of the antibiotic, only cells containing plasmids will survive. Each original bacterial cell that was initially transfected will, in general, contain a different foreign DNA insert; this insert will be inherited by all of the progeny cells of that bacterium, which together form a small colony in a culture dish.

The mixture of many different surviving bacteria contains the DNA library, composed of a large number of different DNA inserts. The problem is that only a few of the bacteria will harbor the particular recombinant plasmids that contain the desired gene. One needs to be able to identify these rare cells in order to recover the DNA of interest in pure form and in useful quantities. Before discussing how this is achieved, we need to describe a second strategy for generating a DNA library that is commonly used in gene cloning.

Two Types of DNA Libraries Serve Different Purposes [17]

Cleaving the entire genome of a cell with a specific restriction nuclease as just described is sometimes called the "shotgun" approach to gene cloning. It produces a very large number of DNA fragments—on the order of a million for a mammalian genome—which will generate millions of different colonies of transfected bacterial cells. Each of these colonies will be composed of a *clone* derived from a single ancestor cell and therefore will harbor a recombinant plasmid with the same inserted genomic DNA sequence. Such a plasmid is said to contain a **genomic DNA clone,** and the entire collection of plasmids is said to constitute

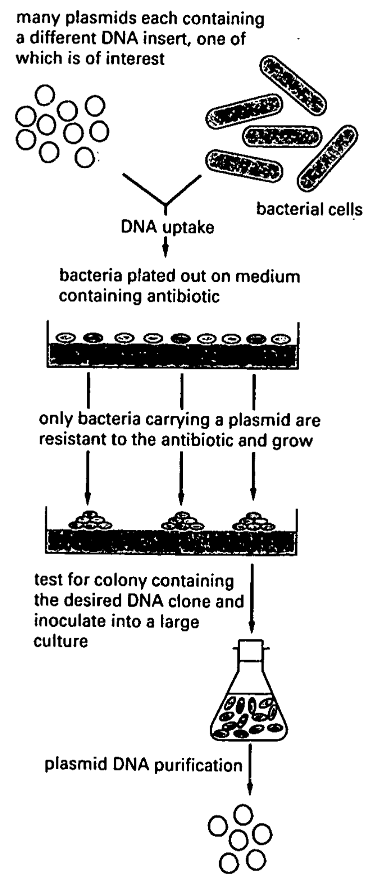


Figure 7–22 Purification and amplification of a specific DNA sequence by DNA cloning in a bacterium. Each bacterial cell carrying a recombinant plasmid develops into a colony of identical cells, visible as a spot on the nutrient agar. By inoculating a single colony of interest into a liquid culture, one can obtain a large number of identical plasmid DNA molecules, each containing the same DNA insert.

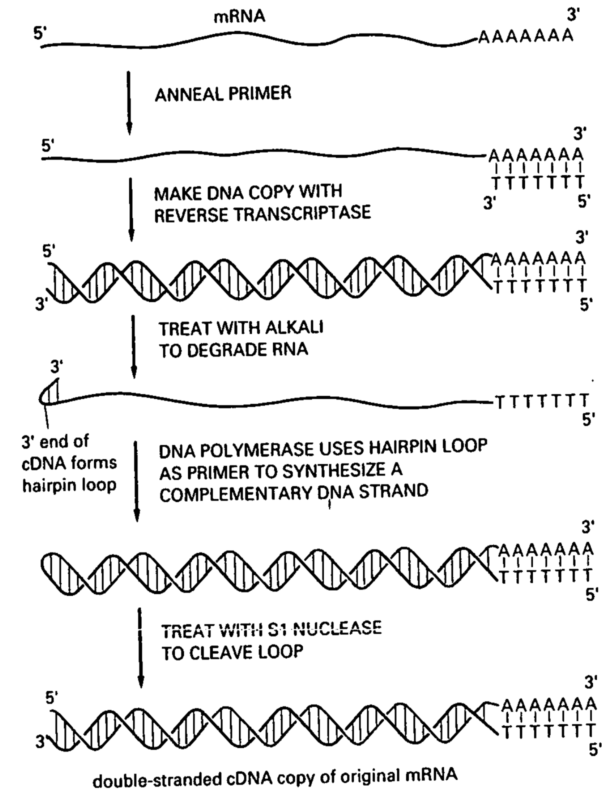


Figure 7–23 The synthesis of cDNA. A DNA copy (cDNA) of an mRNA molecule is produced by the enzyme reverse transcriptase (see p. 282), thereby forming a DNA/RNA hybrid helix. Treating the DNA/RNA hybrid with alkali selectively degrades the RNA strand into nucleotides. The remaining single-stranded cDNA is then copied into double-stranded cDNA by the enzyme DNA polymerase. As indicated, both reverse transcriptase and DNA polymerase require a primer to begin their synthesis. For reverse transcriptase a small oligonucleotide is used; in this example oligo(dT) has been annealed with the long poly-A tract at the 3′ end of most mRNAs. Note that the double-stranded cDNA molecule produced here lacks cohesive ends; such blunt-ended DNA molecules can be cloned by one of several procedures that are analogous to (but less efficient than) that shown in Figure 7–21.

a **genomic DNA library.** But because the genomic DNA is cut into fragments at random, only some fragments will contain genes; many will contain only a portion of a gene, while most of the genomic DNA clones obtained from the DNA of a higher eucaryotic cell will contain only noncoding DNA, which, as we shall discuss in Chapter 8, makes up most of the DNA in such genomes.

An alternative strategy is to begin the cloning process by selecting only those DNA sequences that are transcribed into RNA and thus are presumed to correspond to genes. This is done by extracting the mRNA (or a purified subfraction of the mRNA) from cells and then making a **complementary DNA (cDNA)** copy of each mRNA molecule present; this reaction is catalyzed by the *reverse transcriptase* enzyme of retroviruses, which synthesizes a DNA chain on an RNA template. The single-stranded DNA molecules synthesized by the reverse transcriptase are converted into double-stranded DNA molecules by DNA polymerase, and these molecules are inserted into a plasmid or virus vector and cloned (Figure 7–23). Each clone obtained in this way is called a **cDNA clone,** and the entire collection of clones derived from one mRNA preparation constitutes a **cDNA library.**

There are important differences between genomic DNA clones and cDNA clones, as illustrated in Figure 7–24. Genomic clones represent a random sample of all of the DNA sequences in an organism and, with very rare exceptions, will be the same regardless of the cell type used to prepare them. By contrast, cDNA clones contain only those regions of the genome that have been transcribed into mRNA; as the cells of different tissues produce distinct sets of mRNA molecules, a different cDNA library will be obtained for each type of cell used to prepare the library.

cDNA Clones Contain Uninterrupted Coding Sequences [18]

The use of a cDNA library for gene cloning has several advantages. First, some proteins are produced in very large quantities by specialized cells. In this case,

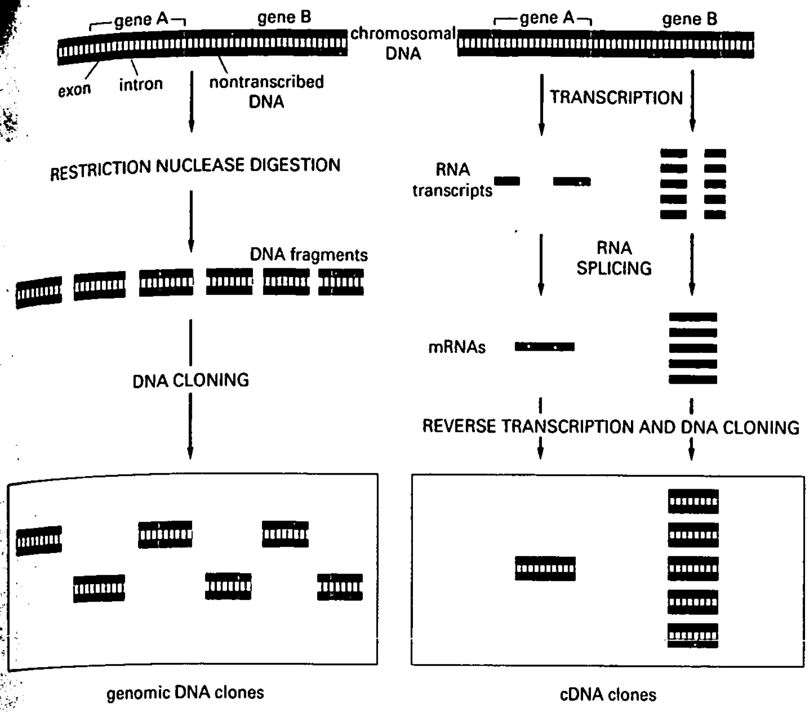


Figure 7–24 The differences between cDNA clones and genomic DNA clones. In this example gene A is infrequently transcribed while gene B is frequently transcribed, and both genes contain introns (*green*). In the genomic DNA clones both the introns and the nontranscribed DNA are included, and most clones will contain only part of the coding sequence of a gene. In the cDNA clones the intron sequences have been removed by RNA splicing during the formation of the mRNA, and a continuous coding sequence is therefore present.

the mRNA encoding the protein is likely to be produced in such large quantities that a cDNA library prepared from the cells will be highly enriched for the cDNA molecules encoding the protein, greatly reducing the problem of identifying the desired clone in the library (see Figure 7–24). Hemoglobin, for example, is made in large amounts by developing erythrocytes (red blood cells); for this reason the globin genes were among the first to be cloned.

By far the most important advantage of cDNA clones is that they contain the uninterrupted coding sequence of a gene. Eucaryotic genes usually consist of short coding sequences of DNA (exons) separated by longer noncoding sequences (introns); the production of mRNA entails the removal of the noncoding sequences from the initial RNA transcript and the splicing together of the coding sequences. Neither bacterial nor yeast cells will make these modifications to the RNA produced from a gene of a higher eucaryotic cell. Thus, if the aim of the cloning is either to deduce the amino acid sequence of the protein from the DNA or to produce the protein in bulk by expressing the cloned gene in a bacterial or yeast cell, it is much preferable to start with cDNA.

Genomic and cDNA libraries are inexhaustible resources that are widely shared among investigators. Today, many such libraries are also available from commercial sources.

cDNA Libraries Can Be Prepared from Selected Populations of mRNA Molecules [19]

When cDNAs are prepared from cells that express the gene of interest at extremely high levels, the majority of cDNA clones may contain the gene sequence, which can therefore be selected with minimal effort. For less abundantly transcribed genes, various methods can be used to enrich for particular mRNAs before making the cDNA library. If an antibody against the protein is available, for example, it can be used to precipitate selectively those polyribosomes (see pp. 237–238) that have the appropriate growing polypeptide chains attached to them. Since these polyribosomes will also have attached to them the mRNA coding for the protein, the precipitate may be enriched in the desired mRNA by as much as 1000-fold.

Subtractiv hybridization provides a powerful alternative way of enriching for particular nucleotide sequences prior to cDNA cloning. This selection procedure can be used, for example, if two closely related cell types are available from the same organism, only one of which produces the protein or proteins of interest. It was first used to identify cell-surface receptor proteins present on T lymphocytes but not on B lymphocytes. It can also be used wherever a cell that expresses the protein has a mutant counterpart that does not. The first step is to synthesize cDNA molecules using the mRNA from the cell type that makes the protein of interest. These cDNAs are then hybridized with a large excess of mRNA molecules from the second cell type. Those rare cDNA sequences that fail to find a complementary mRNA partner are likely to represent mRNA sequences present only in the first cell type. Because these cDNAs remain unpaired after the hybridization, they can be purified by a simple biochemical procedure (a hydroxyapatite column) that separates single-stranded from double-stranded nucleic acids (Figure 7–25). Besides providing a powerful way to clone genes whose products are known to be restricted to a specific differentiated cell type, cDNA libraries prepared after subtractive hybridization are useful for defining the differences in gene expression between any two related types of cells.

Either a DNA Probe or a Test for Expressed Protein Can Be Used to Identify the Clones of Interest in a DNA Library [20]

The most difficult part of gene cloning is often the identification of the rare colonies in the library that contain the DNA fragment of interest. This is especially true in the case of a genomic library, where one has to identify one bacterial cell in a million to select a specific mammalian gene. The technique most frequently used is a form of *in situ* hybridization that takes advantage of the exquisite specificity of the base-pairing interactions between two complementary nucleic acid molecules. Culture dishes containing the growing bacterial colonies are blotted with a piece of filter paper, to which some members of each bacterial colony adhere. The adhering colonies, known as *replicas*, are treated with alkali to disrupt the cells and to separate the strands of their DNA molecules; the paper is then incubated with either a radioactive or a chemically labeled DNA probe containing part of the sequence of the gene being sought (Figure 7–26). If necessary, millions of bacterial clones can be screened in this way to find the one clone that hybridizes with the probe.

In order to find the clone of interest, a specific probe must be made. How this is done will depend on the information that is available about the gene to be cloned. In many cases the protein of interest has been identified by biochemical studies and purified in small amounts. Only a few micrograms of pure protein are often enough to determine the sequence of 30 or so amino acid residues. From this amino acid sequence the corresponding nucleotide sequence can be deduced using the genetic code (with some ambiguities corresponding to amino acids that can be represented by several alternative codons). Two sets of DNA oligonucleotides, chosen to match different parts of the predicted nucleotide sequence of the gene, are then synthesized by chemical methods (Figure 7–27). Colonies of cells that hybridize with both sets of DNA probes are strong candidates for containing the desired gene and are saved for further characterization (see below).

Probes can also be obtained in other ways. If an antibody is available that recognizes the protein produced by the gene, it can be labeled and used as a probe to find a clone that is producing the protein, which therefore contains the desired gene. Any other ligand that is known to bind to the protein encoded by the gene can also be used as a probe: if the gene encodes a receptor protein, for example, the ligand that normally binds to the receptor can, in principle, be used as a probe.

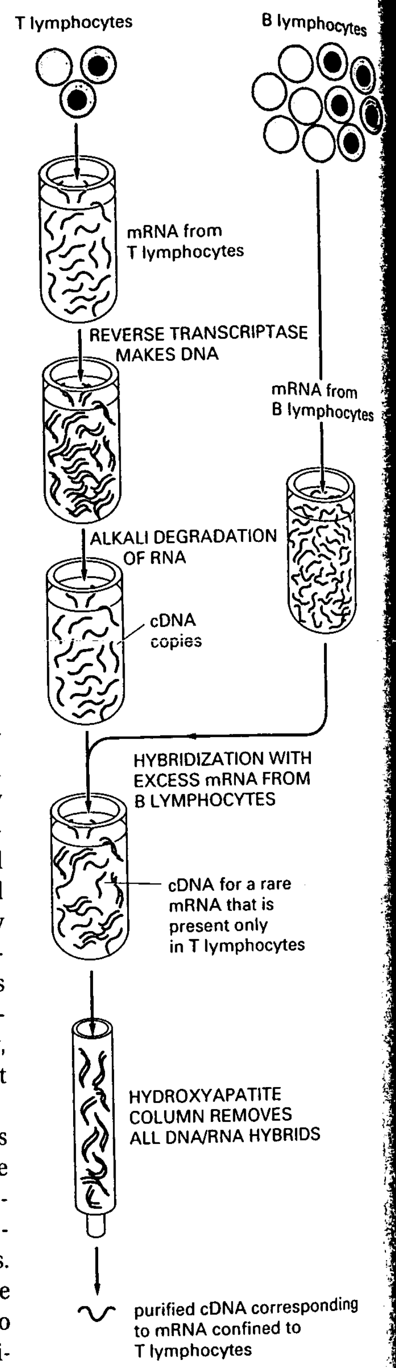


Figure 7–25 Subtractive hybridization. In this example the technique is used to purify rare cDNA clones corresponding to mRNA molecules present in T lymphocytes but not in B lymphocytes. Because the two cell types are very closely related, most of the mRNAs will be common to both cell types; subtractive hybridization is thus a powerful way to enrich for those specialized molecules that distinguish the two cells.

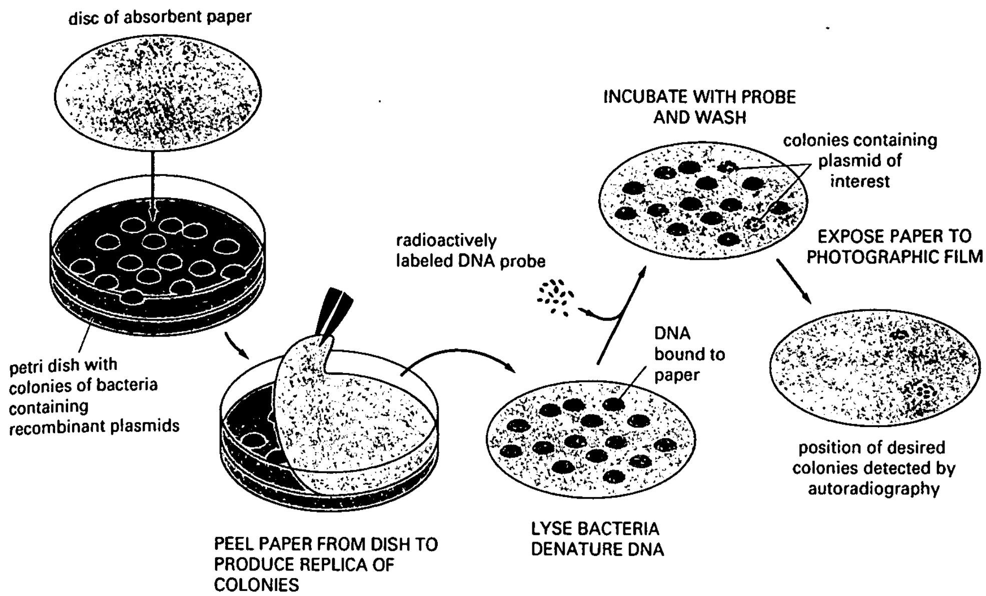


Figure 7–26 An efficient technique commonly used to detect a bacterial colony carrying a particular DNA clone. A replica of the culture is made by pressing a piece of absorbent paper against the surface. This replica is treated with alkali (to disrupt the cells and denature the plasmid DNA) and then hybridized to a highly radioactive DNA probe. Those bacterial colonies that have bound the probe are identified by autoradiography. (See also Figure 7–22.)

Whenever the protein product of a gene is to be detected rather than the gene itself, a special type of cDNA library is required. It is prepared in a special plasmid or virus called an *expression vector*, which directs the transfected bacterium to synthesize large amounts of the protein encoded by the foreign DNA insert contained within the vector's DNA, as we shall discuss later.

In Vitro Translation Facilitates Identification of the Correct DNA Clone [21]

Any method that is used to find a specific clone from a cDNA or genomic DNA library will usually pick out many false positive clones. Further ingenuity is required to discriminate between these and the authentic clones desired. The task is easiest when the desired clone encodes a protein that has already been characterized by other means. In this case each candidate DNA can be tested by one of several methods for its ability to encode the appropriate protein. The cloned DNA can be inserted into an expression vector, for example, so that the protein that it encodes is produced in large amounts in a bacterium. Alternatively, the cloned DNA can be used to obtain a corresponding RNA molecule, either through *in vitro* synthesis with a purified RNA polymerase (see Figure 7–36) or by a technique called *hybrid selection*. In the latter method a mixture of cellular RNAs is added to an excess of single strands of the candidate DNA, and DNA/RNA hybridization is used to purify complementary mRNA molecules from the mixture. In either case the mRNA obtained is allowed to direct protein synthesis in a cell-free system using radioactive amino acids, and the radioactive protein produced is then characterized and compared with the expected protein product of the desired clone. A match in any of these tests allows one to conclude that a cloned DNA fragment encodes the correct protein.

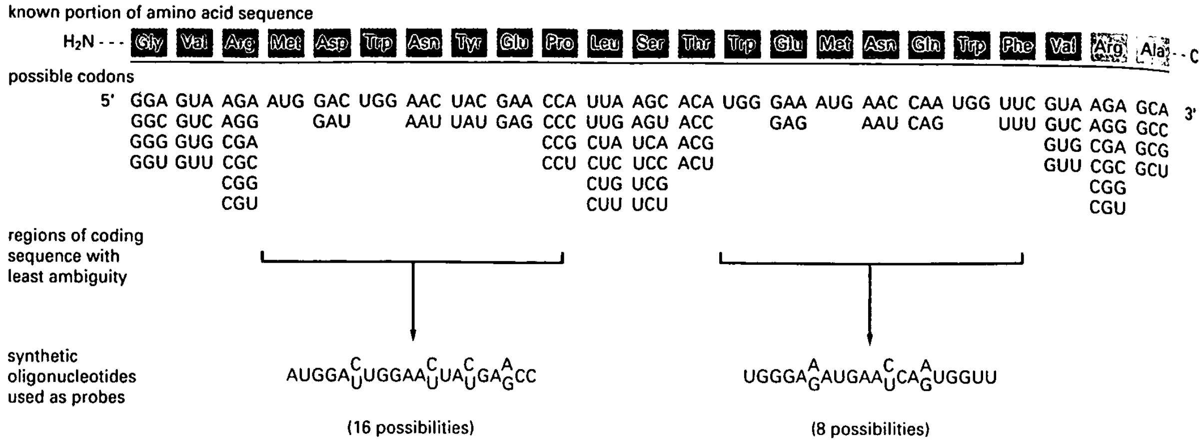


Figure 7–27 Selecting regions of a known amino acid sequence to mak synthetic oligonucleotide probes. Although only one nucleotide sequence will actually code for the protein, the degeneracy of the geneti code means that several different nucleotide sequences will give the same amino acid sequence, and it is impossible to tell in advance which i the correct one. Because it is desirabl to have as large a fraction of the correct nucleotide sequence as possible in the mixture of oligonucleotides to be used as a probe, those regions with the fewest possibilities are chosen, as illustrated In this example the mixture of 8 closely related oligonucleotides shown might be synthesized and use to probe a clone library, and the indicated mixture of 16 oligonucleotides would be used to reprobe all positive clones to find those that actually code for the desired protein. After the oligonucleotide mixture is synthesized by chemical means, the 5 end of each oligonucleotide is radioactively labeled (see Figure 7–6B); alternatively, the probe can be marked with a chemical label by incorporating a modified nucleotide during its synthesis (see Figure 7–18).

The Selection of Overlapping DNA Clones Allows One to "Walk" Along the Chromosome to a Nearby Gene of Interest [22]

Many of the most interesting genes—for example, those that control development—are known only from genetic analysis of mutants in such organisms as the fruit fly *Drosophila* and the nematode *Caenorhabditis elegans.* The protein products of these genes are unknown and may be present in very small quantities in a few cells or produced only at one stage of development. A study of the genetic linkage between different mutations, however, can be used to generate *chromosome maps,* which give the relative locations of the genes (see Figure 7–16). Once one mapped gene has been cloned, the clones in a genomic DNA library that correspond to neighboring genes can be identified using a technique called **chromosome walking.** The methods described in this chapter can then be used to deduce the exact structure and function of the gene of interest and the protein that it encodes.

In chromosome walking one starts with a DNA clone corresponding to a gene or an RFLP marker that is known to be as close as possible to the gene of interest. One end of this clone is used to prepare a DNA probe, which is then used in DNA hybridization experiments to find an overlapping clone in a genomic DNA clone library. The DNA from this second DNA clone is purified, and its far end is used to prepare a second DNA probe, which is used to find a clone that is overlapping, and so on. In this way one can walk along a chromosome one clone at a time, in steps of 30,000 base pairs or more in either direction (Figure 7–28).

How does one know when the gene of interest (identified originally by a deleterious mutation) has been reached, given that the walk is generally too long for complete DNA sequencing to be practicable? For experimental organisms such as fruit flies, nematodes, *Arabidopsis,* yeast, and mice, the ultimate proof of the correct gene is to transfer the normal form of the gene (as a cloned DNA molecule) into a chromosome of the mutant organism, producing a transgenic organism (see Figures 7–45 and 7–49). If the original mutation was a recessive one, the correct DNA should reverse the original mutant phenotype. Other, less stringent criteria, however, are often used and are necessary in the case of human genes, as described later (see Figure 7–30).

Ordered Genomic Clone Libraries Are Being Produced for Selected Organisms [23]

The whole task of identifying mutant genes should become vastly easier as knowledge of the sequence of the normal genome becomes more complete and systematic. By using methods related to those described for chromosome walking, it has been possible to order (map) a complete, or nearly complete, set of large

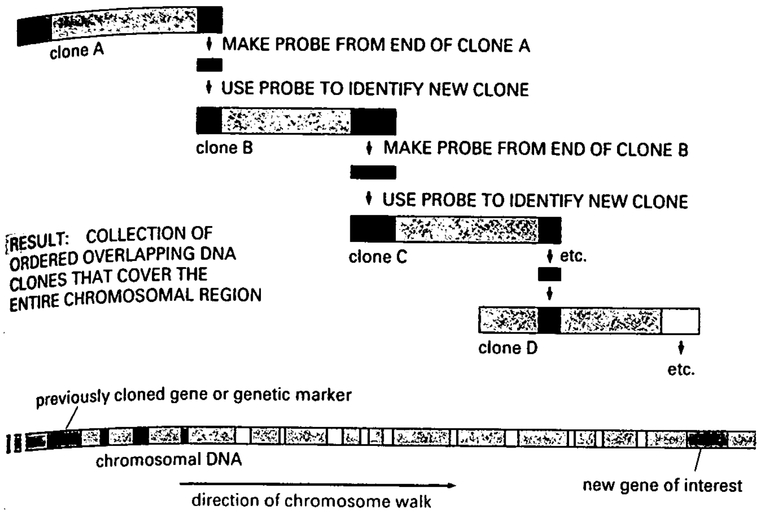


Figure 7–28 The use of overlapping DNA clones to find a new gene by "chromosome walking." To speed up the walk, genomic libraries containing very large cloned DNA molecules are optimal. To probe for the next clone in the walk by DNA hybridization, a short DNA fragment (labeled with a chemical or a radioisotope) from one end of the previously identified clone is purified: If a "right-handed" end is used, for example, the walk will go in the "rightward" direction, as shown in this example. Use of a small end fragment as a probe also reduces the probability that the probe will contain a repeated DNA sequence that would hybridize with many clones from different parts of the genome and thereby interrupt the walk.

genomic clones along the chromosomes of the *E. coli* bacterium, the yeast *Saccharomyces cerevisiae*, the fruit fly *Drosophila*, the plant *Arabidopsis*, and the nematode *C. elegans*. Such large clones, each about 30,000 base pairs in length, are usually prepared in bacteriophage lambda vectors called *cosmids*, which are specially designed to accept only large DNA inserts. It takes a few thousand such clones to cover the entire genome of an organism such as *C. elegans* or *Drosophila*. To map the entire human genome in this way would require ordering more than 100,000 clones in cosmids, which is very time consuming but technically feasible. DNA fragments that are more than 10 times larger than these clones (300,000 to 1.5 million base pairs) can be cloned in yeast cells as *YACs (yeast artificial chromosomes)* (Figure 7–29); in principle, the human genome could be represented as about 10,000 clones of this type (see Figure 8–5).

In the near future, ordered sets of genomic clones will no doubt be available from centralized DNA libraries for use by all research workers. Eventually, a complete library will be available for each commonly studied organism, with each DNA fragment catalogued according to its chromosome of origin and numbered sequentially with respect to the positions of all other DNA fragments derived from the same chromosome. One will then begin a "chromosome walk" simply by obtaining from the library all the clones covering the region of the genome that contains the mutant gene of interest.

Positional DNA Cloning Reveals Human Genes with Unanticipated Functions [24]

Thousands of human diseases are caused by alterations in single genes. Our understanding of these genetic diseases is being revolutionized by recombinant

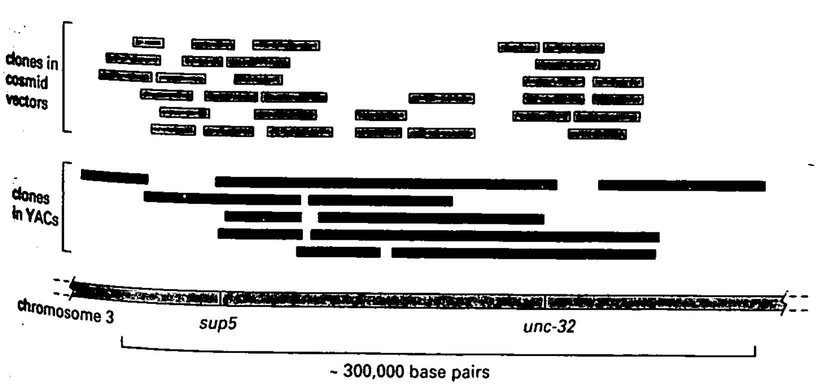


Figure 7–29 Overlapping genomic DNA clones. The collection of clones shown covers a small region of a chromosome of the nematode worm *Caenorhabditis elegans* and represents 0.3% of the total genome. (Adapted from J. Sulston et al., *Nature* 356:37–41, 1992. © 1992 Macmillan Magazines Ltd.)

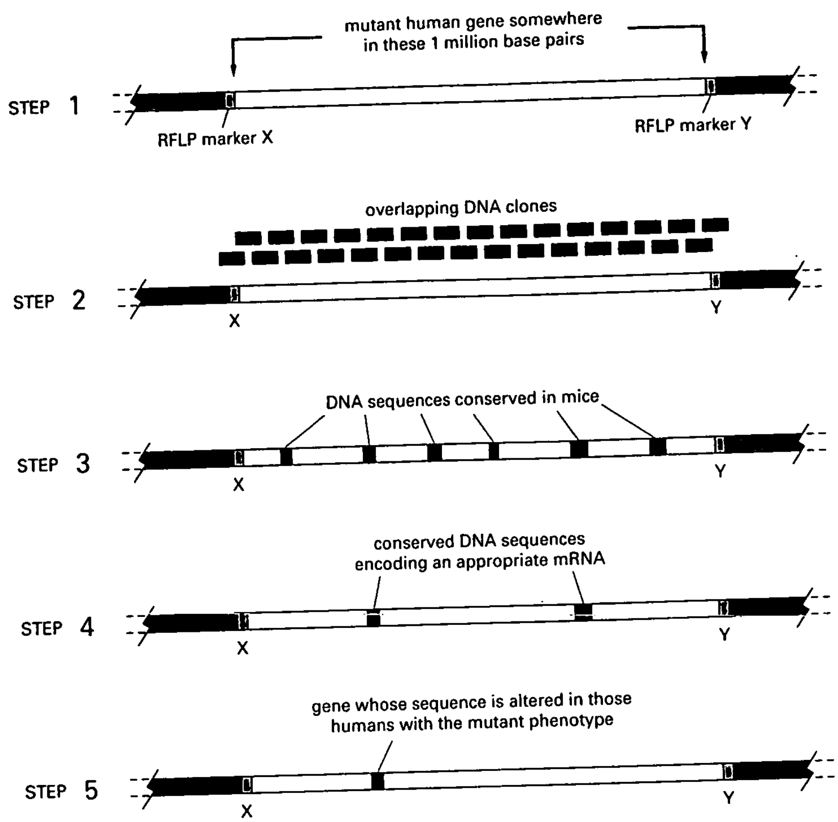


Figure 7–30 Positional cloning. The procedure requires a mutant human gene whose inheritance can be traced in many family groups by virtue of the phenotype that the mutation causes. Step 1, *genetic mapping:* RFLP markers that are coinherited with the phenotype are identified and used to position the gene within about 10^6 base pairs (one megabase, or about 1% the length of a typical human chromosome). Step 2, *assembly of an ordered clone library:* genomic DNA clones are obtained that cover the entire region between two RFLP markers that bracket the gene. Step 3, *search for conserved DNA sequences:* the portions of each DNA clone that hybridize with mouse DNA are identified; only those regions of the human chromosome whose nucleotide sequence is important will have been sufficiently conserved during evolution to form such a hybrid DNA helix (one strand human and the other mouse). Step 4, *search for appropriate mRNAs:* the subset of conserved DNA sequences that encode an mRNA in tissues where the mutant phenotype is expressed are the most likely to represent the mutant gene. Step 5, *finding a difference in the DNA sequence of mutated genes.* When the deleterious mutations in a typical human gene are analyzed, about 1 in 10 turns out to be a deletion that is easily detected as a change in the size of a restriction fragment detected by Southern blotting. For this reason one generally begins by screening the DNA of many human patients with the same disease using probes identified in step 4, looking for such a change. If the mutation is not detectable as a deletion, other, more laborious methods that are capable of detecting single base changes must be used to identify the gene of interest.

DNA methods, which allow the altered DNA to be cloned and sequenced, revealing the precise defect in each patient. In this way, for example, Duchenne's muscular dystrophy was shown to be due to an abnormal cytoskeletal protein in muscle cells, and cystic fibrosis to be an abnormal chloride channel in epithelial cells. This knowledge not only improves the accuracy of diagnosis but also makes it possible, in principle at least, to design treatments.

Although the techniques used to find human disease genes have differed depending on the disease, a standard approach has recently been developed that makes it possible to isolate any human gene that is responsible by itself for a specific trait or disease. It is called **positional cloning** because it starts with genetic linkage mapping to locate the gene in the genome (Figure 7–30). While the approach is straightforward, it presently requires 10 to 100 person-years to isolate a gene in this way. It will become much easier once DNA sequencing is highly automated and the full DNA sequence of the human genome is known: genetic linkage mapping will reveal immediately which genes are prime suspects, and the sequences of these genes can then be analyzed directly in individual patients. The functions of thousands of human genes are likely to be identified in this way.

Selected DNA Segments Can Be Cloned in a Test Tube by a Polymerase Chain Reaction [25]

The availability of purified DNA polymerases and chemically synthesized DNA oligonucleotides has made it possible to clone specific DNA sequences rapidly without the need for a living cell. The technique, called the **polymerase chain reaction (PCR)**, allows the DNA from a selected region of a genome to be amplified a billionfold, provided that at least part of its nucleotide sequence is already known. First, the known part of the sequence is used to design two synthetic DNA oligonucleotides, one complementary to each strand of the DNA double helix and lying on opposite sides of the region to be amplified. These oligonucleotides serve as primers for *in vitro* DNA synthesis, which is catalyzed

by a DNA polymerase, and they determine the ends of the final DNA fragment that is obtained (Figure 7–31).

The principle of the PCR technique is illustrated in Figure 7–32. Each cycle of the reaction requires a brief heat treatment to separate the two strands of the genomic DNA double helix (step 1). The success of the technique depends on the use of a special DNA polymerase isolated from a thermophilic bacterium that is stable at much higher temperatures than normal, so that it is not denatured by the repeated heat treatments. A subsequent cooling of the DNA in the presence of a large excess of the two primer DNA oligonucleotides allows these oligonucleotides to hybridize to complementary sequences in the genomic DNA (step 2). The annealed mixture is then incubated with DNA polymerase and the four deoxyribonucleoside triphosphates so that the regions of DNA downstream from each of the two primers are selectively synthesized (step 3). When the procedure is repeated, the newly synthesized fragments serve as templates in their turn, and within a few cycles the predominant product is a single species of DNA fragment whose length corresponds to the distance between the two original primers. In practice, 20 to 30 cycles of reaction are required for effective DNA amplification. Each cycle doubles the amount of DNA synthesized in the previous cycle. A single cycle requires only about 5 minutes, and an automated procedure permits "cell-free molecular cloning" of a DNA fragment in a few hours, compared with the several days required for standard cloning procedures.

The PCR method is extremely sensitive; it can detect a single DNA molecule in a sample. Trace amounts of RNA can be analyzed in the same way by first transcribing them into DNA with reverse transcriptase. The PCR cloning technique is rapidly replacing Southern blotting for the diagnosis of genetic diseases and for the detection of low levels of viral infection. It also has great promise in forensic medicine as a means of analyzing minute traces of blood or other tissues—even as little as a single cell—and identifying the person from whom they came by his or her genetic "fingerprint" (Figure 7–33).

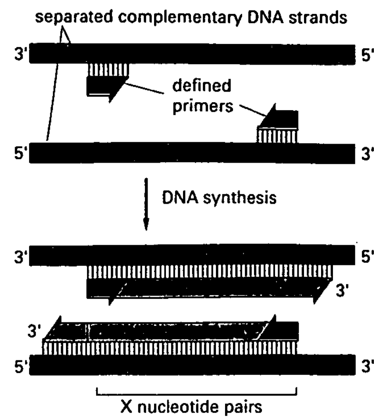
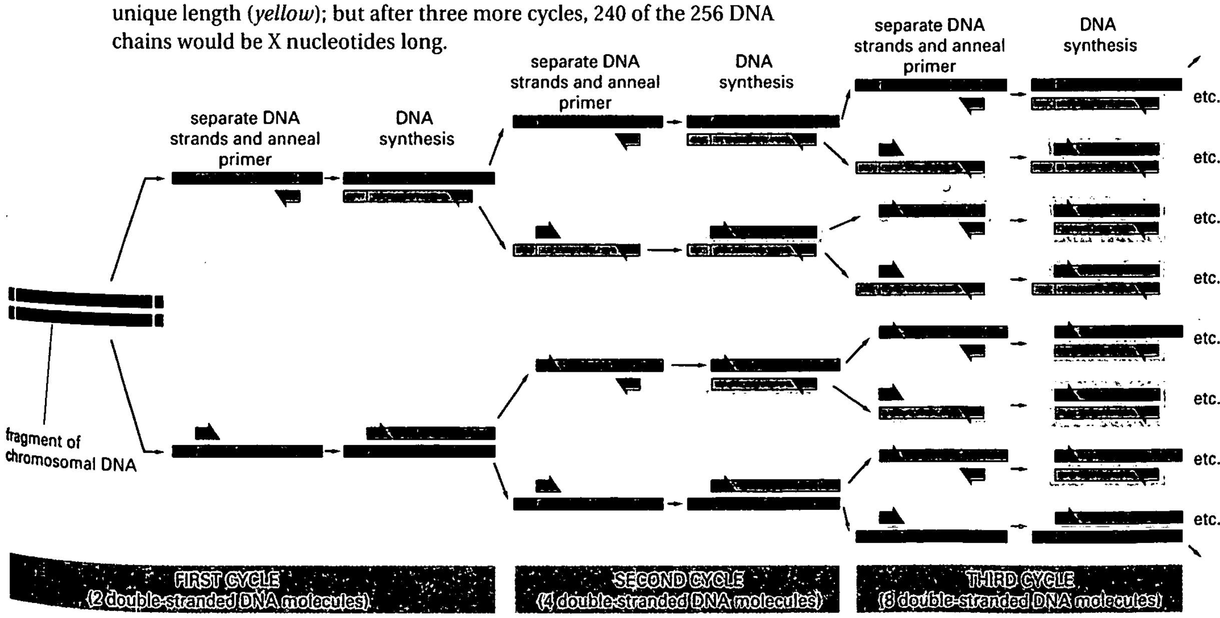


Figure 7–31 The start of the polymerase chain reaction (PCR) for amplifying specific nucleotide sequences *in vitro*. DNA isolated from cells is heated to separate its complementary strands. These strands are then annealed with an excess of two DNA oligonucleotides (each 15 to 20 nucleotides long) that have been chemically synthesized to match sequences separated by X nucleotides (where X is generally between 50 and 2000). The two oligonucleotides serve as specific primers for *in vitro* DNA synthesis catalyzed by DNA polymerase, which copies the DNA between the sequences corresponding to the two oligonucleotides.

Figure 7–32 PCR amplification. PCR produces an amount of DNA that doubles in each cycle of DNA synthesis and includes a uniquely sized DNA species. Three steps constitute each cycle, as described in the text. After many cycles of reaction, the population of DNA molecules becomes dominated by a single DNA fragment, X nucleotides long, provided that the original DNA sample contains the DNA sequence that was anticipated when the two oligonucleotides were designed. In the example illustrated, three cycles of reaction produce 16 DNA chains, 8 of which have this unique length (*yellow*); but after three more cycles, 240 of the 256 DNA chains would be X nucleotides long.



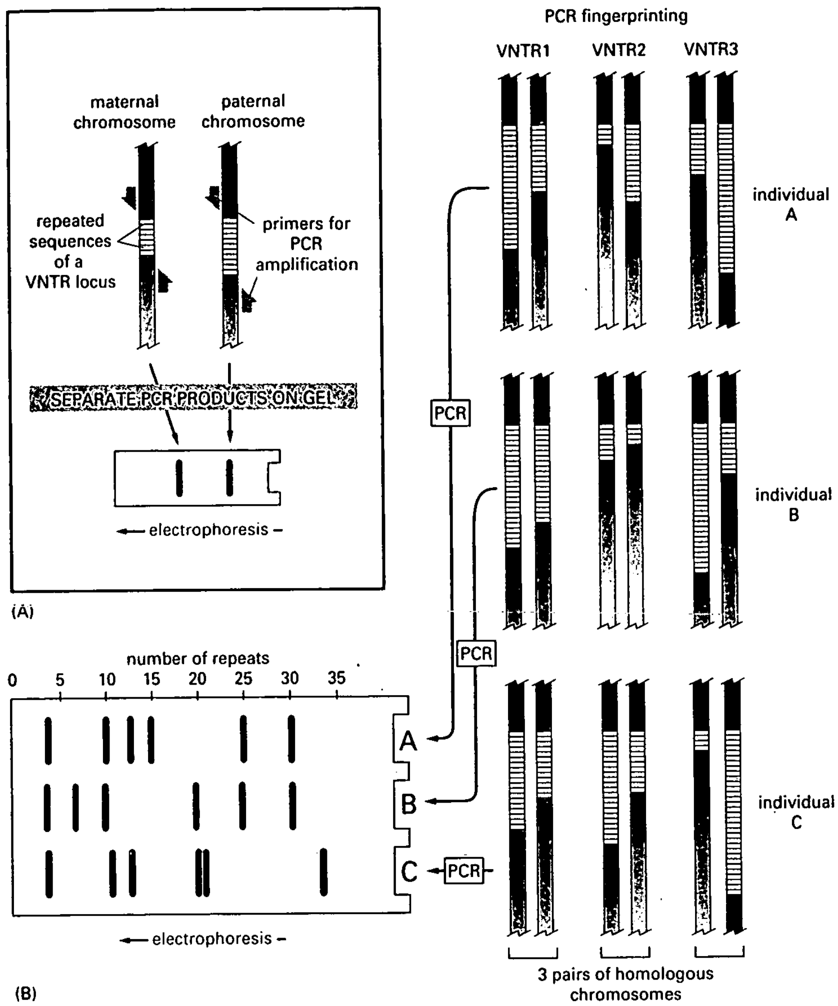


Figure 7–33 The use of PCR in forensic science. (A) A PCR reaction using two primers that bracket a particular microsatellite, or VNTR, sequence (see Figure 7–14C) produces a different pair of DNA bands from each individual. One of these bands contains the repeated VNTR sequence that was inherited from the individual's mother and the other contains the repeated VNTR sequence that was inherited from the individual's father. (B) The large set of DNA bands obtained from a set of different PCR reactions, each of which amplifies the DNA from a different VNTR sequence, can serve as a "fingerprint" to identify each individual nearly uniquely. The starting material for the PCR reaction can be a single hair that was left at the scene of a crime.

Summary

DNA cloning allows a copy of any specific part of a DNA or RNA sequence to be selected from the millions of other sequences in a cell and produced in unlimited amounts in pure form. DNA sequences are amplified after cutting chromosomal DNA with a restriction nuclease and inserting the resulting DNA fragments into the chromosome of a self-replicating genetic element (a plasmid or a virus). When a plasmid vector is used, the resulting "genomic DNA library" is housed in millions of bacterial cells, each carrying a different cloned DNA fragment. The bacterial colony containing a DNA fragment of interest is identified by hybridization using a DNA probe or, following expression of a cloned gene or gene fragment in the bacterial host cell, by using a test that detects the desired protein product. The cells in the identified bacterial colony are then allowed to proliferate, producing large amounts of the desired DNA fragment.

The procedure used to obtain DNA clones that correspond in sequence to mRNA molecules are the same except that the starting material is a DNA copy of the mRNA sequence, called cDNA, rather than fragments of chromosomal DNA. Unlike genomic DNA clones, cDNA clones lack intron sequences, making them the clones of choice for expressing and characterizing the protein product of a gene.

PCR is a new form of DNA cloning that is carried out outside cells using a purified, thermostable DNA polymerase enzyme. This type of DNA amplification requires a prior knowledge of gene sequence, since two synthetic oligonucleotide primers must be synthesized that bracket the DNA sequence to be amplified. PCR cloning, however, has the advantage of being much faster and easier than standard cloning methods.